5.3 Project: Big Data Project Using Hadoop Hive

3. a)

**Create and Load Hive Table:**

CREATE TABLE IF NOT EXISTS default.`2009\_Jan\_Sales` (

Transaction\_date STRING,

Product STRING,

Price INT,

Payment\_Type STRING,

Name STRING,

City STRING,

State STRING,

Country STRING,

Account\_Created STRING,

Last\_Login STRING,

Latitude DOUBLE,

Longitude DOUBLE

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

LOAD DATA INPATH '/user/hive/warehouse/SalesJan2009.csv' INTO TABLE default.`2009\_Jan\_Sales`;

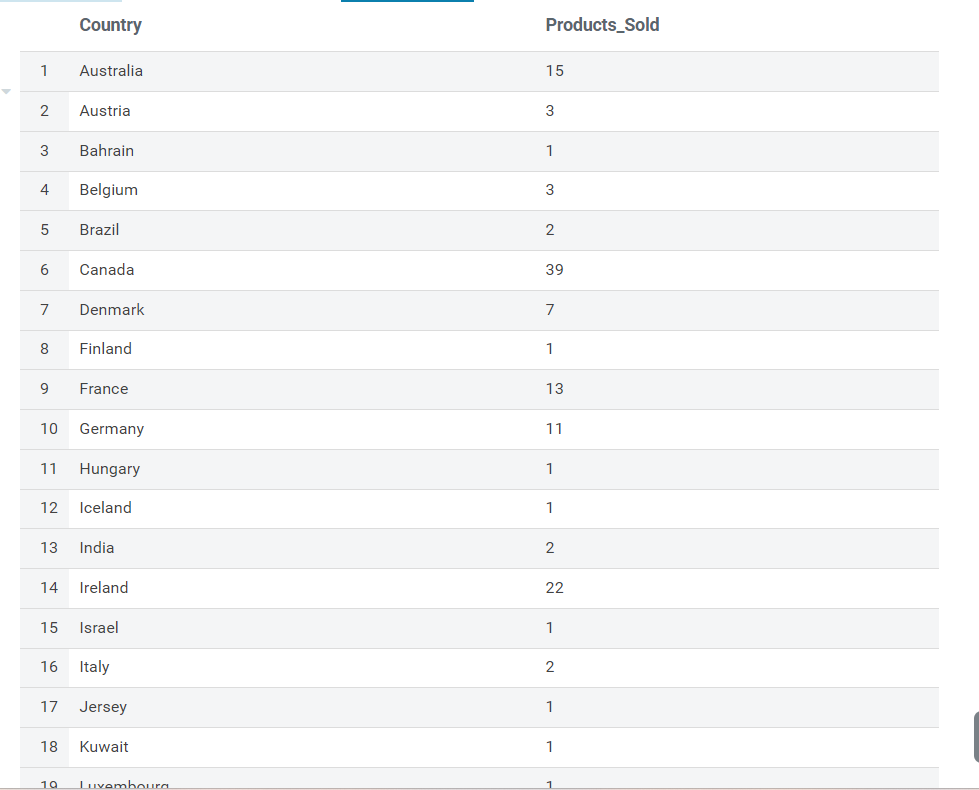
Query : Total number of products sold in each country

SELECT Country, COUNT(\*) AS Products\_Sold

FROM default.`2009\_Jan\_Sales`

GROUP BY Country;

Results :



Exercise 3b )

* Average Price per Payment Type
* Number of Distinct Customers per Country
* Total Number of Transactions per Day

1. **Average Price per Payment Type**

SELECT Payment\_Type, AVG(Price) AS Average\_Price

FROM default.`2009\_Jan\_Sales`

GROUP BY Payment\_Type;

Result :



**ii) Number of Distinct Customers per Country**

SELECT Country, COUNT(DISTINCT Name) AS Distinct\_Customers

FROM default.`2009\_Jan\_Sales`

GROUP BY Country LIMIT 10

